

## REPORT DOCUMENTATION PAGE

AD-A222 852

|  |  |   |                          |
|--|--|---|--------------------------|
| 2b. DECLASSIFICATION/DOWNGRADING SCHEDULE  |  | 1b. RESTRICTIVE MARKINGS  |                          |
| 4. PERFORMING ORGANIZATION REPORT NUMBER(S)<br>City College, CUNY No. MB89-03  |  | 3. DISTRIBUTION/AVAILABILITY OF REPORT<br>Approved for public release; distribution unlimited.                          |                          |
| 6a. NAME OF PERFORMING ORGANIZATION<br>City College, CUNY  | 6b. OFFICE SYMBOL<br>(If applicable)       | 5. MONITORING ORGANIZATION REPORT NUMBER(S)<br>AFOSR 89-03<br><b>AFOSR-TR-89-0580</b>                                   |                          |
| 7a. NAME OF MONITORING ORGANIZATION<br>Air Force Office of Scientific Research   |  | 7b. ADDRESS (City, State and ZIP Code)<br>Directorate of Mathematical & Information Sciences, Bolling AFB DC 20332-6448 |                          |
| 8a. NAME OF FUNDING/SPONSORING ORGANIZATION<br>AFOSR   | 8b. OFFICE SYMBOL<br>(If applicable)<br>NM | 9. PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER<br><b>AFOSR 89-0083</b>   |                          |
| 10. SOURCE OF FUNDING NOS.   |  | 11. TITLE (Include Security Classification)<br>Consequences of Monotonicity for Markov Transition Functions             |                          |
| 10a. PROGRAM ELEMENT NO.<br>61102F   |  | 10b. PROJECT NO.<br>2304  | 10c. TASK NO.<br>A5      |
| 10d. WORK UNIT NO.   |  | 12. PERSONAL AUTHOR(S)<br>Mark Brown  |                          |
| 13a. TYPE OF REPORT<br>Technical   |  | 13b. TIME COVERED<br>FROM 89 TO 1990  |                          |
| 14. DATE OF REPORT (Yr., Mo., Day)<br>February 28, 1990  |  | 15. PAGE COUNT<br>45  |                          |
| 16. SUPPLEMENTARY NOTATION   |  |   |                          |
| 17. COSATI CODES   |  | 18. SUBJECT TERMS (Continue on reverse if necessary and identify by block number)                                       |                          |
| FIELD  | GROUP                                      | SUB. GR.  |                          |
|  |  |   |                          |
|  |  |   |                          |
| 19. ABSTRACT (Continue on reverse if necessary and identify by block number)   |  |   |                          |
| <p>For an ergodic Markov chain we examine conditions for, and consequences of, monotonicity in time of <math>P_{ij}(t) = \sum \pi_0(i) \Pr(X(t)=j   X(0)=i)</math>, for some choice of <math>\pi_0</math> and <math>j</math>. The work has application to the study of asymptotic exponentiality of first passage time distributions, as well to the computation of separation distance.</p> |  |   |                          |
| 20. DISTRIBUTION/AVAILABILITY OF ABSTRACT<br>UNCLASSIFIED/UNLIMITED <input checked="" type="checkbox"/> SAME AS RPT. <input type="checkbox"/> DTIC USERS <input type="checkbox"/>  |  | 21. ABSTRACT SECURITY CLASSIFICATION<br>UNCLASSIFIED  |                          |
| 22a. NAME OF RESPONSIBLE INDIVIDUAL<br>Dr. Jon H. Sorenson   |  | 22b. TELEPHONE NUMBER<br>(Include Area Code)<br>(202) 767-4940  | 22c. OFFICE SYMBOL<br>NM |

Consequences of Monotonicity for  
Markov Transition Functions

By

Mark Brown  
The City College, CUNY

February 1990

|                    |                                     |
|--------------------|-------------------------------------|
| Accession For      |                                     |
| NTIS GRA&I         | <input checked="" type="checkbox"/> |
| DTIC TAB           | <input type="checkbox"/>            |
| Unannounced        | <input type="checkbox"/>            |
| Justification      |                                     |
| By                 |                                     |
| Distribution/      |                                     |
| Availability Codes |                                     |
| Dist               | Special                             |
| A-1                |                                     |

City College, CUNY Report No. MB89-03  
AFOSR Report No. 89-03  
AFOSR Grant No. 89-0083

90 05 25 188

### Section 1. Introduction and Summary.

For an ergodic Markov chain in discrete or continuous time, we will be interested in situations in which there exists an initial distribution  $\pi_0$ , and a state  $j$ , such that  $\pi_t(j) = P_{\pi_0}(X(t)=j)$  is increasing in  $t$ . Also of interest are states,  $j$ , for which  $p_t(j,j) = \Pr(X(t)=j|X(0)=j)$  is decreasing in  $t$ . We will examine the implications of such monotonicity, as well as conditions for it to hold.

Let  $\pi$  denote the stationary distribution and  $T_{\pi,j}$  the waiting time, starting in steady state, to reach state  $j$ , with  $T_{\pi,j} = 0$  if  $X(0) = j$ . Similarly define  $T_{\pi_0,j}$ . In Section 2 it is shown that if  $\pi_t(j) = P_{\pi_0}(X(t)=j)$  is increasing, then:

$$(1.1) \quad T_{\pi_0,j} \sim Y + T_{\pi,j}$$

where  $\Pr(Y > t) = 1 - \frac{\pi_t(j)}{\pi(j)}$ , and  $Y$  and  $T_{\pi,j}$  are independent. One consequence of (1.1) is that  $T_{\pi_0,j}$  is stochastically larger than  $T_{\pi,j}$ ; another is that if  $T_{\pi,j}$  is approximately exponential and  $EY$  is small compared to  $ET_{\pi,j}$ , then  $T_{\pi_0,j}$  is also approximately exponential. Approximate exponentiality is discussed in Section 7.

A tempting heuristic interpretation of (1.1) is that  $Y$  represents "the waiting time from  $\pi_0$  to steady state." That is, if there were a random variable  $Y$  with  $X(Y) \sim \pi$ ,  $X(Y)$  independent of  $Y$ , and  $\Pr(X(t)=j, \text{ for some } t < Y) = 0$ , then  $T_{\pi_0,j}$  would equal  $Y$  (the waiting time to steady state) plus  $T_{\pi,j}$  (the waiting time from steady state to  $j$ ), and  $Y$  and

$T_{\pi,j}$  would be independent. This suggests that the distribution of  $Y$  (given above) is that of a strong stationary time, in the sense of Aldous and Diaconis (1987). This interpretation holds precisely in a large class of situations described in Section 4. However, in general,  $\pi_t(j)$  increasing does not imply that the distribution of  $Y$  is that of a strong stationary time.

If  $p_t(j,j)$  is decreasing, in discrete or continuous time, for a state  $j$ , then  $T_{\pi,j}$  can be represented as a geometric convolution. Specifically:

$$(1.2) \quad T_{\pi,j} = \sum_{i=1}^N W_i$$

where  $\{W_i, i=1,2,\dots\}$  are i.i.d.,  $N$  is independent of  $\{W_i\}$  with  $\Pr(N=k) = (1-\pi(j))^k \pi(j)$ ,  $k=0,1,\dots$ , and

$$(1.3) \quad \Pr(W > t) = \frac{p_t(j,j) - \pi(j)}{1 - \pi(j)}.$$

This representation has two uses. Firstly, geometric convolutions with small  $p$  (in this case  $p=\pi(j)$ ) are approximately exponential. Error bounds can be obtained from the first two moments of  $T_{\pi,j}$ . Secondly, the moments of  $T_{\pi,j}$  are easily related to those of  $W$ , which in turn can be expressed in terms of the eigenvalues. In the case of time reversible chains in continuous time this leads to:

$$(1.4) \quad \sup |\Pr(T_{\pi,j} > t) - e^{-t/\alpha_j}| \leq \frac{\tau/\alpha_j}{(\tau/\alpha_j) + 1}$$

where  $\alpha_j = ET_{\pi,j}$  and  $\tau$  is the relaxation time, defined as  $\lambda_1^{-1}$ , where  $0 = \lambda_0 > -\lambda_1 \geq -\lambda_2 \geq \dots \geq -\lambda_m$  are the eigenvalues of the infinitesimal matrix. This provides a quantification and generalization of Proposition 7 of Aldous (1989), who showed that  $\tau/\alpha_j$  was a key parameter in approximate exponentiality for random walks on vertex transitive graphs.

We now describe the class of situations for  $\pi_t(j)$  monotonicity previously alluded to. Consider ergodic Markov chains in discrete or continuous time, taking values in a partially ordered set  $S$ , possessing a unique maximum state  $M$  ( $i \leq M$  for all  $i \in S$ ). Further, assume that the time reversed process is stochastically monotone relative to the partial ordering, and that  $\pi_0(k)/\pi(k)$  is decreasing in  $k \in S$  relative to the partial ordering. Then  $\pi_t(M)$  is increasing in  $t$ , and:

$$(1.5) \quad s(t) = \max_k \left( 1 - \frac{\pi_t(k)}{\pi(k)} \right) = 1 - \frac{\pi_t(M)}{\pi(M)}.$$

The quantity  $s(t)$  is the separation between  $\pi_t$  and  $\pi$ , as studied by Aldous and Diaconis (1987), and Diaconis and Fill (1989). It yields an upper bound for total variation distance, which is the main quantity of interest in the study of how rapidly a Markov chain approaches ergodicity. Expression (1.5) tells us that for all  $t$ , the separation is achieved at state  $M$ , which we call a separating state. This greatly simplifies the computation of  $s(t)$  and examples are presented to illustrate this point.

In the case where  $S$  is totally ordered, the above conditions coincide with those of Theorem 4.6 of Diaconis and Fill (1989), under which a dual process with convenient properties is constructed, and then employed to study  $s(t)$ . The current approach offers an alternative to duality, and allows for the flexibility of choice of partial ordering of  $S$ . Examples are given to show that for the same Markov chain, different partial orderings can be used for different initial conditions, resulting in a wide array of  $\pi_t(j)$  monotonicity and convenient computation of  $s(t)$ .

Another class of processes considered are discrete time ergodic Markov chains with state space  $\{0, \dots, M\}$ , satisfying  $\bar{P}(i_1, j_1)\bar{P}(i_2, j_2) \geq \bar{P}(i_1, j_2)\bar{P}(i_2, j_1)$  for all  $i_1 < i_2, j_1 < j_2$ , where  $\bar{P}(i, j) = \sum_j^m P(i, k)$ . For such a chain, if  $\pi_0(i)/\pi(i)$  is decreasing, then  $\pi_n(M)$  is increasing in  $n$ , and (1.5) holds.

Section 2. Derivation of (1.1).

Consider, first, a continuous time chain. Let  $T_{\pi,j}$  denote the first passage time to  $j$  under  $X_0 \sim \pi$ , with  $T_{\pi,j} = 0$  if  $X(0) = j$ ; similarly define  $T_{\pi_0,j}$ . Now:

$$(2.1) \quad \pi_j = P_{\pi}(X(t)=j) = \int_0^t p_{t-x}^{(j,j)} dF_{\pi,j}^{(x)}$$

where  $F_{\pi,j}$  is the distribution of  $T_{\pi,j}$ . Take Laplace transforms in (2.1) to obtain:

$$(2.2) \quad \psi_{\pi,j}(s) = \int e^{-st} dF_{\pi,j}(t) = \frac{\pi_j}{s\psi_{jj}(s)}$$

where  $\psi_{jj}$  is the Laplace transform of  $p_t(j,j)$ .

Assume that  $\pi_t(j) = P_{\pi_0}(X(t)=j)$  is increasing in  $t$ . Since  $\lim_{t \rightarrow \infty} \pi_t(j) = \pi(j)$ ,  $\pi_t(j) \leq \pi(j)$  for all  $t$ , thus  $\pi_t(j)/\pi(j)$  is a cdf. Let  $Y$  denote a random variable with this cdf, and let  $\psi_Y$  denote the Laplace transform of  $Y$ ,  $\psi_t$  the Laplace transform of  $\pi_t(j)$ , and  $\psi_{\pi_0,j}$  the Laplace transform of  $T_{\pi_0,j}$ . Then:

$$(2.3) \quad \psi_Y(s) = s \int_0^{\infty} e^{-st} \Pr(Y \leq t) dt = \frac{s}{\pi(j)} \psi_t(s) .$$

Analogous to (2.1) we have:

$$(2.4) \quad \pi_t(j) = \int_0^t p_{t-x}^{(j,j)} dF_{\pi_0,j}^{(x)} .$$

From (2.1), (2.3) and (2.4):

$$(2.5) \quad \psi_{\pi_0, j}^{(s)} = \frac{\psi_t^{(s)}}{\psi_{jj}^{(s)}} = \left( \frac{\pi_j}{s \psi_{jj}^{(s)}} \right) \left( \frac{s}{\pi_j} \psi_t^{(s)} \right) = \psi_{\pi, j}^{(s)} \psi_Y^{(s)}.$$

Thus  $T_{\pi_0, j} \sim Y + T_{\pi, j}$ , and (1.1) is proved.

In the discrete time case  $Y$  is a discrete random variable with cdf

$F(n) = \frac{\pi_n(j)}{\pi(j)}$ . The above argument holds using probability generating functions in place of Laplace transforms.

### Section 3. Conditions for Monotonicity of $\pi_t(j)$ .

First we present some elementary facts about stochastically monotone Markov chains, with finite partially ordered state spaces.

Let  $S$  be a finite set with partial ordering denoted by  $\leq^*$ . Define  $A$  to be an upper set if  $x \in A$  and  $y \geq^* x$  implies  $y \in A$ . Similarly, define lower sets. Define a discrete time Markov chain with state space  $S$  and probability transition matrix  $P$  to be stochastically monotone relative to  $\leq^*$  if:

$$P(x, A) = \sum_{y \in A} P(x, y) \leq P(y, A)$$

for all  $x \leq^* y$  and upper sets  $A$ . Define a real-valued function  $h$  on  $S$  to be increasing relative to  $\leq^*$  if  $x \leq^* y$  implies  $h(x) \leq h(y)$ . If  $h$  is increasing then the elements of  $S$  can be labeled as  $d_1, \dots, d_n$ , in such a way that  $i < j$  implies  $h(d_i) \leq h(d_j)$ , and that  $d_i$  is not greater than  $d_j$  under the partial ordering (Brown and Chaganty (1983) p. 1007). It follows that  $A_j = \{d_j, d_{j+1}, \dots, d_n\}$  is an upper set,  $j=1, \dots, n$ . Defining  $h(d_0) = 0$  ( $d_0$  is not in  $S$ ) we have:

$$\begin{aligned} E_i h(X) &= E(h(X_1) | X_0 = i) = \sum_j P(i, j) h(j) \\ &= \sum_{j=1}^n P(i, d_j) \sum_{r=1}^{d_j} (h(d_r) - h(d_{r-1})) = \sum_{r=1}^n (h(d_r) - h(d_{r-1})) P(i, A_r) . \end{aligned}$$

It follows that if  $P$  is stochastically monotone and  $h$  is increasing, then  $E_i h(X)$  is increasing (all relative to  $\leq^*$ ). Since:

$$p_{n+1}(i, A) = \Pr(X_{n+1} \in A | X_0 = i) = E_i p_n(X, A)$$

it follows by induction that stochastic monotonicity implies that  $p_n(i, A)$  is increasing in  $i$ , for all  $n$  and upper sets  $A$ . Similarly  $E_i h(X_n)$  is increasing in  $i$  for all increasing  $h$ , and all  $n$ .

In the continuous time case, let  $\{X(t), t \geq 0\}$  be a Markov chain with state space  $S$  and infinitesimal matrix  $Q$ . Define  $\{X(t), t \geq 0\}$  to be stochastically monotone, relative to  $\leq^*$ , if both of the following hold:

$$(i) \quad Q(x, A) = \sum_{y \in A} q(x, y) \leq Q(y, A)$$

for all  $x \leq^* y$ , and upper sets  $A$  not containing  $y$ .

$$(ii) \quad Q(x, B) \geq Q(y, B) \text{ for all } x \leq^* y,$$

and lower sets  $B$  not containing  $x$ .

If (i) and (ii) hold, choose  $c \geq 2 \max_i \sum_{k \neq i} q_{ik}$ , and define  $P = I + c^{-1}Q$ . Then  $P$  is stochastically monotone as is seen in the three possible cases:

1) If  $x <^* y$  and  $A$  is an upper set not containing  $y$  then:

$$P(x, A) = c^{-1}Q(x, A) \leq c^{-1}Q(y, A) = P(y, A).$$

2) If  $x <^* y$  and  $A$  is an upper set containing  $x$ :

$$P(x, A) = 1 - c^{-1}Q(x, \tilde{A}) \leq 1 - c^{-1}Q(y, \tilde{A}) = P(y, A)$$

where  $\tilde{A}$ , the complement of  $A$ , is a lower set.

3) If  $x \stackrel{*}{<} y$  and  $A$  is an upper set containing  $y$  but not  $x$ , then:

$$P(x, A) = c^{-1} Q(x, A)$$

$$P(y, A) = 1 - c^{-1} Q(y, \tilde{A})$$

thus,

$$P(y, A) - P(x, A) = 1 - c^{-1} [Q(x, A) + Q(y, \tilde{A})] \geq 1 - \frac{Q(x, A) + Q(y, \tilde{A})}{(2 \max_i \sum_{k \neq i} q_{ik})} \geq 0.$$

Using stochastic monotonicity of  $P$ , we find that for  $x \stackrel{*}{\leq} y$ , and upper sets  $A$ :

$$p_t(x, A) = \sum \frac{(ct)^n e^{-ct}}{n!} p_n(x, A) \leq \sum \frac{(ct)^n e^{-ct}}{n!} p_n(y, A) = p_t(y, A).$$

Thus  $P_i(X(t) \in A)$  is increasing in  $i$  for all  $t$  and upper sets  $A$ . Similarly if  $h$  is increasing then  $E_i(h(X(t)))$  is increasing in  $i$  for all  $t$ . We summarize the above:

Lemma 3.1. Let  $\{X_t, t=0, 1, \dots\}$  or  $\{X(t), t \geq 0\}$  be a Markov chain with finite state space  $S$ , stochastically monotone relative to  $\stackrel{*}{\leq}$ . Then:

- (i)  $p_t(x, A)$  is increasing in  $x$  for all  $t$  and upper sets  $A$ .
- (ii)  $E_i h(X_t)$  is increasing in  $i$  for all  $t$  and increasing  $h$ .  $\square$

If  $\{X_n, n=0, 1, \dots\}$  is ergodic with stationary distribution  $\pi$ , the time reversed chain is the Markov chain with transition matrix  $\tilde{P}(i, j) = \frac{\pi(j)}{\pi(i)} P(j, i)$ . Similarly for continuous time ergodic chains, the reversed chain has infinitesimal matrix,  $\tilde{Q}(i, j) = \frac{\pi_j}{\pi_i} Q(j, i)$ .

Time reversible chains satisfy  $P = \tilde{P}$ , in discrete time, and  $Q = \tilde{Q}$  in continuous time.

Theorem 3.2, below, derives conditions for  $\pi_t(j)$  increasing in  $t$ .

Theorem 3.2. Let  $\{X_t, t=0,1,\dots\}$  or  $\{X(t), t \geq 0\}$  be an ergodic Markov chain taking values in  $S$ , where  $S$  is a finite partially ordered set with partial ordering  $\leq^*$ . Assume that  $S$  contains a unique maximal element  $M$ , so that  $x \leq^* M$  for all  $x \in S$ . Then if the time reversed process,  $\tilde{X}$ , is stochastically monotone, and  $\pi_0(x)/\pi(x)$  is decreasing in  $x \in S$  (both with respect to  $\leq^*$ ), then  $\pi_t(M)$  is increasing in  $t$ .

Proof. We first show that  $\pi_t(M)/\pi(M)$  minimizes  $\pi_t(j)/\pi(j)$ , over  $j \in S$ .

This follows since:

$$\begin{aligned} (3.3) \quad \frac{\pi_t(M)}{\pi(M)} &= \sum \frac{\pi_0(i)}{\pi(M)} p_t(i, M) = \sum \frac{\pi_0(i)}{\pi(i)} \tilde{p}_t(M, i) \\ &= E_M \left[ \frac{\pi_0}{\pi} (\tilde{X}_t) \right] \leq E_j \left[ \frac{\pi_0}{\pi} (\tilde{X}_t) \right] = \frac{\pi_t(j)}{\pi(j)} \end{aligned}$$

the last inequality holding since  $\frac{\pi_0}{\pi}$  is decreasing and  $\tilde{X}$  is stochastically monotone.

To show that (3.3) implies that  $\pi_t(M)$  is increasing in  $t$ , consider the hypothesis testing problem,  $H_0: X_0 \sim \pi_0$  vs  $H_1: X_0 \sim \pi$ , based on the data  $(X_s, X_t)$ . The densities under  $H_0$  and  $H_1$  are given by  $f_0(x, y) = \pi_s(x) p_{t-s}^{(x, y)}$  and  $f_1(x, y) = \pi(x) p_{t-s}^{(x, y)}$ . The test which has smallest type 1 error among all tests with type 2 error  $\leq 1 - \pi(M)$ , rejects  $H_0$  if  $X(s) = M$  (Neyman-Pearson Lemma). The type 1 error of this test is  $\pi_s(M)$ . A less efficient competing test with type 2 error  $1 - \pi(M)$ , rejects  $H_0$  if  $X(t) = M$ , and has larger type 1 error,  $\pi_t(M)$ . Thus  $\pi_s(M) \leq \pi_t(M)$  for  $s \leq t$ .  $\square$

#### Section 4. Separation Distance and Strong Stationary Times.

Separation distance and its connection to strong stationary times is an elegant contribution of Aldous and Diaconis (1987), with important recent developments by Diaconis and Fill (1989).

For an ergodic Markov chain with finite state space, and initial distribution  $\pi_0$ , the separation at  $t$  is defined by:

$$s(t) = \max_k \left( 1 - \frac{\pi_t(k)}{\pi(k)} \right).$$

Separation provides an upper bound for the total variation norm:

$$\begin{aligned} (4.1) \quad d(t) &= \max_{B \subseteq S} |\pi_t(B) - \pi(B)| = \sum_{\pi(k) > \pi_t(k)} (\pi(k) - \pi_t(k)) \\ &= E_{\pi} \left( 1 - \frac{\pi_t}{\pi}(X) \right)^+ \leq P_{\pi}(\pi(X) > \pi_t(X)) s(t) \leq s(t). \end{aligned}$$

Call the Markov chain separable with separating state  $M$ , under  $\pi_0$ , if  $s(t) = 1 - \frac{\pi_t(M)}{\pi(M)}$  for all  $t$ .

A strong stationary time is a stopping time,  $T$ , with  $X(T) \sim \pi$  and  $X(T)$  independent of  $T$ . For any strong stationary time,  $s(t) \geq \Pr(T > t)$  (Aldous and Diaconis (1987) p. 72). When equality holds for all  $t$ ,  $T$  is called a minimal strong stationary time. Aldous and Diaconis (1987) construct a minimal strong stationary time for a general ergodic, finite state, discrete time Markov chain.

Corollary 4.1. Under the conditions of Theorem 3.2 the Markov chain is separable with separating state  $M$ . Furthermore there exists a minimal strong stationary time,  $Y$ , satisfying:

$$(i) \quad \Pr(Y > t) = 1 - \frac{\pi_t(M)}{\pi(M)} = s(t), \text{ for all } t.$$

$$(ii) \quad \Pr(X(t)=M, \text{ for some } t < Y) = 0.$$

Define  $Z = [\min\{t \geq Y : X(t) = j\} - Y]$ . Then  $Z$  is independent of  $Y$  and distributed as  $T_{\pi, M}$ . Thus  $T_{\pi_0, M} = Y + Z$  with  $Y$  and  $Z$  independent,  $\Pr(Y > t) = 1 - \frac{\pi_t(M)}{\pi(M)} = s(t)$ , and  $Z \sim T_{\pi, M}$ .

Proof. By the proof of Theorem 3.2:

$$\frac{\pi_t(M)}{\pi(M)} \leq \frac{\pi_t(j)}{\pi(j)} \text{ for all } j \in S$$

thus  $s(t) = 1 - \frac{\pi_t(M)}{\pi(M)}$ , and the Markov chain is separable with separating state  $M$ .

In discrete time, the Aldous-Diaconis construction produces a minimal strong stationary time,  $Y$ , which by the details of their construction satisfies (ii). Since  $Y$  is a minimal strong stationary time, and the process is separable,  $\Pr(Y > n) = s(n) = 1 - \frac{\pi_n(M)}{\pi(M)}$ .

For a continuous time chain, by choosing  $c = 2 \max_{i \neq k} \sum q_{ik}$ , we obtain a discrete time skeleton,  $P = I + c^{-1}Q$ , which satisfies the above conditions. Denoting the embedded Markov chain by  $\{X'_n, n=0,1,\dots\}$ , we can represent  $\{X(t), t \geq 0\}$  by  $\{X(t) = X'_{N(t)}, t \geq 0\}$  where  $\{N(t), t \geq 0\}$  is a Poisson process of rate  $c$ , independent of  $\{X'_n, n=0,1,\dots\}$ . Denote the event epochs from the Poisson process as  $\{S_n, n \geq 1\}$ , and define:

$$Y = S_{Y'},$$

where  $Y'$  is a minimal strong stationary time for  $\{X'_n\}$ .

Clearly,  $Y$  is a strong stationary time for the continuous time process and:

$$\begin{aligned} \Pr(Y > t) &= \sum_{n=0}^{\infty} \frac{(ct)^n e^{-ct}}{n!} \Pr(Y' > n) = \sum_{n=0}^{\infty} \frac{(ct)^n e^{-ct}}{n!} \left(1 - \frac{\pi_n(M)}{\pi(M)}\right) \\ &= 1 - \frac{\pi_t(M)}{\pi(M)} = s(t) . \end{aligned}$$

Thus  $Y$  is a minimal strong stationary time. Furthermore:

$$\Pr(X(t)=M, \text{ for some } t < Y) = \Pr(X'_n=M, \text{ for some } n < Y') = 0 .$$

Thus (i) and (ii) hold in continuous time.

In view of (ii),  $T_{\pi_0, M} = Y+Z$ . Define  $\{X^*(t), t=0, 1, \dots\}$  ( $\{X^*(t), t \geq 0\}$  in continuous time) by  $X^*(t) = X(Y+t)$ . Since  $Y$  is a strong stationary time,  $\{X^*(t)\}$  is independent of  $Y$ . Since  $Z$  is the first passage time to  $M$  for the  $X^*$  process,  $Z$  is also independent of  $Y$ . Since  $X^*(0) = X(Y) \sim \pi$ , it follows that  $Z \sim T_{\pi, M}$ . Thus  $T_{\pi_0, M} = Y+Z$ , with  $Y$  and  $Z$  independent,  $\Pr(Y > t) = s(t)$ , and  $Z \sim T_{\pi, M}$ .

(4.2) Remark. Define the chain distance from state  $x$  to state  $y$ ,  $d(x, y)$ , to be the minimal  $n$  such that  $p_n(x, y) > 0$ . If  $\pi_0 = \delta_x$  (one point distribution at  $\{x\}$ ) then a necessary condition for  $M$  to be a separating state is that  $d(x, y) \leq d(x, M)$  for all  $y \in S$ . This is true since if  $M$  is a separating state and  $p_k(x, M) > 0$ , then:

$$1 > s(k) = 1 - \frac{p_k(x, M)}{\pi(M)} \geq 1 - \frac{p_k(x, y)}{\pi(y)}$$

thus  $p_k(x, y) > 0$ .

In applications there is often only one potential separating state. The problem then reduces to producing a partial ordering under which  $M$  is the unique maximal state,  $x$  is a minimal state, and  $\tilde{P}$  is stochastically monotone. In section 5, this point is illustrated by examples.  $\square$

Under the conditions of Corollary 4.1,  $\pi_t(j)/\pi(j)$  is decreasing in  $j$  and thus assumes its minimum at  $M$ . Of course monotonicity of  $\pi_t(j)/\pi(j)$  is considerably stronger than  $\pi_t(j)/\pi(j) \geq \pi_t(M)/\pi(M)$ . For a totally ordered state space, say  $\{0, \dots, M\}$  with  $0 < 1 < 2 < \dots < M$ , a weaker condition than  $\pi_t(j)/\pi(j)$  decreasing, is  $\bar{\pi}_t(j)/\bar{\pi}(j)$  decreasing, where  $\bar{\pi}_t(j) = \sum_{k=j}^M \pi_t(k)$ ,  $\bar{\pi}(j) = \sum_{k=j}^M \pi(k)$ . This is true since:

$$(4.2) \quad \frac{\bar{\pi}_t(j)}{\bar{\pi}(j)} = \sum_{k=j}^M \left( \frac{\pi_t(k)}{\pi(k)} \right) \frac{\pi(k)}{\bar{\pi}(j)} = E_{\pi} \left[ \frac{\pi_t}{\pi}(X) \mid X \geq j \right].$$

Since  $\pi_t/\pi$  is decreasing and  $X \mid X \geq j$  stochastically increasing (where  $X \sim \pi$ ), it follows that  $\bar{\pi}_t(j)/\bar{\pi}(j)$  is decreasing.

Furthermore,  $\bar{\pi}_t(j)/\bar{\pi}(j)$  decreasing implies  $\pi_t(j)/\pi(j) \geq \pi_t(M)/\pi(M)$  for  $j = 0, \dots, M$ . To see this define  $y_j = \min\{\frac{\pi_t(k)}{\pi(k)}, k \geq j\}$ ,  $j = 0, \dots, M$ . Note that by (4.2), for  $j=0, \dots, M-1$ :

$$(4.3) \quad \begin{aligned} \frac{\bar{\pi}_t(j)}{\bar{\pi}(j)} &= \left( \frac{\pi(j)}{\bar{\pi}(j)} \right) \left( \frac{\pi_t(j)}{\pi(j)} \right) + \left( 1 - \frac{\pi(j)}{\bar{\pi}(j)} \right) \frac{\bar{\pi}_t(j+1)}{\bar{\pi}(j+1)} \\ &= \left( \frac{\pi(j)}{\bar{\pi}(j)} \right) \left( \frac{\pi_t(j)}{\pi(j)} \right) + \left( 1 - \frac{\pi(j)}{\bar{\pi}(j)} \right) E_{\pi} \left[ \frac{\pi_t}{\pi}(X) \mid X \geq j+1 \right]. \end{aligned}$$

Since  $\bar{\pi}_t/\bar{\pi}$  is decreasing, it follows from (4.3) that:

$$\frac{\pi_t(j)}{\pi(j)} \geq E_{\pi} \left[ \frac{\pi_t(X)}{\pi(X)} | X \geq j+1 \right] \geq y_{j+1}, \quad j=0, \dots, M-1.$$

Thus

$$y_j = \min \left\{ \frac{\pi_t(k)}{\pi(k)}, k \geq j \right\} = \min \left\{ \frac{\pi_t(j)}{\pi(j)}, y_{j+1} \right\} = y_{j+1}$$

for  $j=0, \dots, M-1$ . Thus:

$$(4.4) \quad y_0 = \min \left\{ \frac{\pi_t(k)}{\pi(k)}, k=0, \dots, M \right\} = y_M = \frac{\pi_t(M)}{\pi(M)}.$$

It follows from (4.4) that if for all  $t$ ,  $\bar{\pi}_t/\bar{\pi}$  is decreasing, then the process is separable with separating state  $M$ , and:

$$s(t) = 1 - \frac{\pi_t(M)}{\pi(M)}.$$

We now need to find conditions under which  $\bar{\pi}_t/\bar{\pi}$  is decreasing for all  $t$ .

Define a discrete time Markov chain on  $\{0, \dots, M\}$  to be failure rate monotone if  $i_1 < i_2, j_1 < j_2$  implies:

$$(4.5) \quad \bar{P}(i_1, j_1) \bar{P}(i_2, j_2) \geq \bar{P}(i_1, j_2) \bar{P}(i_2, j_1)$$

where  $\bar{P}(i, j) = \sum_{k=j}^M P(i, k)$ . This condition is equivalent to  $i_1 < i_2$  implies

$\bar{P}(i_1, j)/\bar{P}(i_2, j)$  decreasing in  $j$ , for  $j$  such that  $\bar{P}(i_2, j) > 0$ .

It is also equivalent to  $P(i, j)/\bar{P}(i, j)$  decreasing in  $i$  for each  $j$ , where  $P(i, j)/\bar{P}(i, j)$  is defined to be 1 if  $\bar{P}(i, j) = 0$ . Shantikumar [(1988), Lemma 2.1, p. 399] proves that if (4.5) is satisfied for  $P$ , then it is also satisfied for  $P^n$  for  $n \geq 2$ .

Gathering together the above observations we derive:

Lemma 4.2. Let  $\{X_n, n=0, 1, \dots\}$  be an ergodic Markov chain with state space  $\{0, 1, \dots, M\}$ , which is failure rate monotone. Then  $\pi_0(j)/\pi(j)$  decreasing implies that  $\pi_n(M)$  is increasing in  $n$ , and that  $s(n) = 1 - \frac{\pi_n(M)}{\pi(M)}$ , for all  $n$ .

Proof. It follows by the above remarks that we just need prove that  $\pi_n(j)/\pi(j)$  is decreasing in  $j$ , equivalently that  $\pi_n(j)/\pi_n(j) \leq \pi(j)/\pi(j)$  for all  $j$ . Define

$$h_{n,i}^{(j)} = \frac{p_n(i, j)}{\bar{p}_n(i, j)} \quad \text{and} \quad h_{\pi}(k) = \frac{\pi(k)}{\bar{\pi}(k)}, \quad \text{then}$$

$$h_n(j) = \frac{\pi_n(j)}{\bar{\pi}_n(j)} = \frac{\sum \pi_0(i) \bar{p}_n(i, j) h_{n,i}^{(j)}}{\sum \pi_0(i) \bar{p}_n(i, j)} = \sum c_{n,i}^{(j)} h_{n,i}^{(j)}$$

$$h_{\pi}(j) = \frac{\sum \pi(i) \bar{p}_n(i, j) h_{n,i}^{(j)}}{\sum \pi(i) \bar{p}_n(i, j)} = \sum d_{n,i}^{(j)} h_{n,i}^{(j)}$$

where  $c_{n,i}$  and  $d_{n,i}$  are probability distributions on  $\{0, \dots, M\}$ . Now by failure rate monotonicity of  $P$ ,  $h_{n,i}^{(j)}$  is decreasing in  $i$ . Moreover

$c_{n,i}^{(j)}/d_{n,i}^{(j)} = \alpha(\pi_0(i)/\pi(i))$  with  $\alpha$  a constant, thus  $d_{n,i}$  is larger than  $c_{n,i}$  under monotone likelihood ratio ordering, and thus under stochastic ordering. Thus  $h_n(j) \geq h_\pi(j)$  so that  $\bar{\pi}_n(j)/\bar{\pi}(j)$  is decreasing, and therefore (by (4.4)),  $s(n) = 1 - \frac{\pi_n(M)}{\pi(M)}$ . Since, by the Aldous-Diaconis construction,  $s(n)$  is decreasing, it follows that  $\pi_n(M)$  is increasing.  $\square$

Example. Consider the following Markov chain on  $\{0,1,2\}$

$$P = \begin{pmatrix} 1/4 & 1/2 & 1/4 \\ 1/8 & 1/8 & 3/4 \\ 1/8 & 1/8 & 3/4 \end{pmatrix}.$$

The time reversed chain is given by :

$$\tilde{P} = \begin{pmatrix} 1/4 & 5/32 & 19/32 \\ 2/5 & 1/8 & 19/40 \\ 1/19 & 15/76 & 3/4 \end{pmatrix}.$$

Since  $\tilde{P}(1,0) > \tilde{P}(0,0)$ ,  $\tilde{P}$  is not stochastically monotone, and Corollary 4.1 does not apply. However,  $P$  is failure rate monotone ( $h_0(0) = \frac{1}{4} > \frac{1}{8} = h_1(0) = h_2(0), h_0(1) = \frac{2}{3} > \frac{1}{7} = h_1(1) = h_2(1), h_0(2) = h_1(2) = h_2(2) = 1$ ). We conclude that  $p_n(0,2)$  is increasing, and that  $s_0(n) = 1 - \frac{p_n(0,2)}{\pi(2)} = \frac{96}{19}(\frac{1}{8})^n$ ,  $n \geq 1$ . We further remark that  $\frac{\pi_n(j)}{\pi(j)}$  is not decreasing in  $j$ :

$$\frac{p_n(0,0)}{\pi(0)} = 1 + 6(\frac{1}{8})^n < \frac{p_n(0,1)}{\pi(1)} = 1 + \frac{72}{5}(\frac{1}{8})^n > \frac{p_n(0,2)}{\pi(2)} = 1 - \frac{96}{19}(\frac{1}{8})^n, \quad n \geq 1$$

I have not worked on extending Lemma 4.2 to continuous time, or to partially ordered spaces.

## Section 5. Examples.

### 5.1. Non-symmetric walk on the unit cube.

Consider a discrete time Markov chain,  $\{X_n, n=0,1,\dots\}$ . with state space,  $\{0,1\}^d$ , the  $2^d$  d-tuples of 0's and 1's. The chain evolves by changing at most one coordinate at a time. Thus if  $\delta_i$  is a d-vector with  $\delta_i(j) = \begin{cases} 1 & j=i \\ 0 & j \neq i \end{cases}$  then:

$$P(x, x+\delta_i) = \alpha_i(x) \text{ for } x_i = 0$$

$$P(x, x-\delta_i) = \beta_i(x) \text{ for } x_i = 1$$

$$P(x, x) = 1 - \sum_{x_i=0} \alpha_i(x) - \sum_{x_i=1} \beta_i(x) .$$

Assume that  $\alpha_i(x), \beta_i(x)$  are such that the chain is ergodic. Let  $\pi_0 = \delta_x$ , a point distribution at  $x$ . By Remark (4.2), the only potential separating state is  $x+1$ , where  $(x+1)_i = x_i + 1 \pmod{2}$ . This suggests partial ordering by chain distance (see Remark (4.2)), i.e.  $y \leq^* z$  if and only  $d(x,y) < d(x,z)$ . Under this ordering,  $x$  is the unique minimal state, and  $x+1$  the unique maximal state. The problem then reduces to finding conditions for stochastic monotonicity of  $\tilde{P}$ .

Consider a special case of the above with  $\alpha_i(x) = \alpha_i, \beta_i(x) = \beta_i$ , independent of  $x$ , with:

$$(i) \quad \alpha_i > 0, \beta_i > 0, i=1,\dots,d$$

$$(ii) \quad \sum_A \alpha_i + \sum_{\bar{A}} \beta_i < 1 \text{ for some } A \subset \{1,2,\dots,d\} .$$

$$(iii) \quad \max \left( \sum_{i=1}^d \alpha_i + \max \beta_i, \sum_{i=1}^d \beta_i + \max \alpha_i \right) \leq 1.$$

Condition (i) is necessary and sufficient for irreducibility, and (ii) for aperiodicity. The chain is then easily shown to be time reversible. Condition (iii) insures that the chain is stochastically monotone, with respect to the above partial ordering.

It then follows from Corollary 4.1 that:

$$s_x(n) = \max_y \left( 1 - \frac{p_n(x,y)}{\pi(y)} \right) = 1 - \frac{p_n(x,x+1)}{\pi(x+1)}.$$

To compute  $s_x(n)$ , consider the continuous time chain with  $Q = I - P$ . The continuous time process is composed of  $d$  independent 0-1 processes. Thus:

$$(5.1) \quad p_t(x, x+1) = \left[ \prod_{x_i=0} \frac{\alpha_i}{\alpha_i + \beta_i} (1 - e^{-(\alpha_i + \beta_i)t}) \right] \left[ \prod_{x_i=1} \frac{\beta_i}{\alpha_i + \beta_i} (1 - e^{-(\alpha_i + \beta_i)t}) \right]$$

$$= \pi(x+1) \prod_{i=1}^d (1 - e^{-(\alpha_i + \beta_i)t}) = \pi(x+1) \sum_{k=0}^n (-1)^k \sum_{\gamma \in A_k} e^{-s_\gamma t}$$

where  $A_k$  consists of the  $\binom{d}{k}$  subsets of size  $k$  from  $\{1, \dots, d\}$ , and  $s_\gamma = \sum_{i \in \gamma} (\alpha_i + \beta_i)$ , for  $\gamma$  a subset of  $\{1, \dots, d\}$ . From the above expression for  $p_t(x, x+1)$ , and the spectral representation,  $p_t(x, x+1) = \pi(x+1) + \sqrt{\frac{\pi(x+1)}{\pi(x)}} \sum \mu_j(x) \mu_j(x+1) e^{-\lambda_j t}$  (Keilson (1979) p. 34), we see that the eigenvalues of  $Q$  are  $\{-s_\gamma, \gamma \subset \{1, \dots, d\}\}$  and thus the eigenvalues of

$P = I - Q$  are  $\{1 - s_\gamma, \gamma \in \{1, \dots, d\}\}$ . From the spectral representation, since

$p_n(x, x+1) = \pi(x+1) + \sqrt{\frac{\pi(x+1)}{\pi(x)}} \sum \mu_j(x) \mu_j(x+1) (1 - \lambda_j)^n$  where  $\mu_j(x), \mu_j(x+1)$  are the same terms as in  $p_t(x, x+1)$ , it follows from (5.1) that:

$$(5.2) \quad s_x(n) = 1 - \frac{p_n(x, x+1)}{\pi(x+1)} = \sum_{k=1}^n (-1)^{k-1} \sum_{\gamma \in A_k} (1 - s_\gamma)^n.$$

In the case  $\sum_1^d (\alpha_i + \beta_i) \leq 1$ , we recognize (5.2) as an inclusion-exclusion formula. Specifically, consider  $n$  multinomial trials with cell probabilities,  $p_i = \alpha_i + \beta_i$ ,  $i=1, \dots, d$ , and  $p_{d+1} = 1 - \sum_1^d (\alpha_i + \beta_i)$ . Let  $A_n$  be the event that at least one of the cells  $1, \dots, d$  are empty. Then  $A_n = \bigcup_1^d C_i$ , where  $C_i = \{\text{cell } i \text{ is empty}\}$ , and (5.2) represents the inclusion-exclusion formula for  $\Pr(\bigcup_1^d C_i)$ . Thus if we let  $T$  denote the waiting time for all of cells  $1, \dots, d$  to each be occupied, then:

$$\Pr(T > n) = \Pr(A_n) = s_x(n).$$

By using a construction of the author (Brown (1975), p. 379, (1984) p. 608) we can construct a minimal strong stationary time for the continuous time process, which when applied to the embedded discrete time process is indeed the waiting time for cells  $1, \dots, d$  to be occupied. Adapting the construction to discrete time, requires  $\sum_1^d (\alpha_i + \beta_i) \leq 1$ . When  $\sum_1^d (\alpha_i + \beta_i) > 1$  (with (i), (ii), (iii) still holding), I have no simplifying explanation for (5.2). It is pleasantly surprising that  $s_x(n)$  is independent of  $x$ .

The symmetric walk  $\alpha_i = \beta_i \equiv \frac{1-r}{d}$  has been considered by Diaconis (1988), and Diaconis and Fill (1989). In this case (i) is satisfied for  $r < 1$ , (ii) for  $r > 0$ , and (iii) for  $r \geq \frac{1}{n+1}$ . Diaconis and Fill utilize the symmetry to reduce the problem to a birth and death process on  $\{0, \dots, d\}$ . We discuss birth and death processes in Section 5.2.

### 5.2. Skip free to the right Markov chains.

A Markov chain on  $\{0, \dots, d\}$  is defined to be skip free to the right if  $P(i, j) = 0$  for  $j - i \geq 2$ . An important special case is a birth and death process, which is skip free to both the left and right, i.e.  $P(i, j) = 0$  for  $|j - i| \geq 1$ .

A birth and death chain on  $\{0, \dots, d\}$  is irreducible if  $P(i, i+1)P(i+1, i) > 0$  for  $i=0, \dots, d-1$ . It is aperiodic if  $P(i, i) > 0$  for some  $i$ . An ergodic birth and death chain is time reversible. It is stochastically monotone if and only if  $P(i, i+1) + P(i+1, i) \leq 1$  for  $i=0, \dots, d-1$ . An ergodic birth and death process in continuous time is necessarily stochastically monotone.

It follows from Corollary 4.1 that if a skip free to the right Markov chain is ergodic and  $\tilde{P}$  is stochastically monotone, then under  $\pi_0 = \delta_0$  the chain is separable with separating state  $d$  thus:

$$(5.3) \quad s_0(n) = 1 - \frac{p_n(0, d)}{\pi(d)}.$$

In the Appendix we show that for an ergodic skip free to the right chain on  $[0, d]$ , with distinct eigenvalues  $1, \beta_1, \dots, \beta_d$  that;

$$(5.4) \quad p_n(0,d) = \pi(d) \left[ 1 - \sum_{j=1}^d \beta_j^n \left( \prod_{r \neq j} \frac{1-\beta_r}{\beta_j-\beta_r} \right) \right].$$

From (5.3) and (5.4) we derive:

Lemma 5.1. (i) Let  $\{X_n, n=0,1,\dots\}$  be an ergodic skip free to the right Markov chain on  $[0,\dots,d]$ , with  $P$  processing distinct eigenvalues  $1, \beta_1, \dots, \beta_d$ . Assume that  $\tilde{P}$  is stochastically monotone. Then:

$$s_0(n) = \sum_{j=1}^d \beta_j^n \left( \prod_{r \neq j} \frac{1-\beta_r}{\beta_j-\beta_r} \right).$$

In particular this will hold for an ergodic stochastically monotone birth and death chain on  $[0,\dots,d]$ . When  $\beta_i > 0$ ,  $i=1,\dots,d$ ,  $s_0(n) = \Pr(T > n)$ , where  $T$  is the convolution of  $d$  independent geometric distributions with parameters  $\beta_1, \dots, \beta_d$ .

(ii) Let  $\{X(t), t \geq 0\}$  be an ergodic skip free to the right Markov chain on  $[0,\dots,d]$  with  $Q$  processing distinct eigenvalues  $0, -\lambda_1, \dots, -\lambda_d$ . Assume that  $\tilde{Q}$  is stochastically monotone. Then:

$$s_0(t) = \sum_{j=1}^d \left( \prod_{r \neq j} \frac{\lambda_j}{\lambda_j - \lambda_r} \right) e^{-\lambda_j t}.$$

In particular this will hold for an ergodic birth and death process on  $[0,\dots,d]$ . When  $\lambda_j > 0$ ,  $j=1,\dots,n$ ,  $s_0(t) = \Pr(T > t)$ , where  $T$  is the convolution of  $d$  independent exponential distributions with parameters  $\lambda_1, \dots, \lambda_d$ . This is always the case for birth and death chains.

Proof. (i) The expression for  $s_0(n)$  follows from (5.3) and (5.4). It holds for ergodic stochastically monotone birth and death chains, because such chains necessarily have distinct eigenvalues (Keilson (1979)p. 59). The convolution result follows from an argument, given for the continuous time case, in Brown and Shao (1987), p. 72.

(ii) For  $c$  sufficiently large, the discrete time embedded chain,  $P = I + c^{-1}Q$ , will satisfy the conditions in (i). The eigenvalues of  $P$  are 1,  $\beta_i = 1 - c^{-1}\lambda_i, i=1, \dots, d$ . Note that,  $(1 - \beta_r)/(\beta_j - \beta_r) = \lambda_r/(\lambda_r - \lambda_j)$ . Thus:

$$\begin{aligned} s_0(t) &= \sum_{n=0}^{\infty} \frac{(ct)^n e^{-ct}}{n!} \sum_{j=1}^k \left( \prod_{r \neq j} \frac{\lambda_r}{\lambda_r - \lambda_j} \right) \beta_j^n = \sum_{j=1}^k \left( \prod_{r \neq j} \frac{\lambda_r}{\lambda_r - \lambda_j} \right) e^{-ct(1 - \beta_j)} \\ &= \sum_{j=1}^k \left( \prod_{r \neq j} \left( \frac{\lambda_r}{\lambda_r - \lambda_j} \right) \right) e^{-\lambda_j t}. \end{aligned}$$

The convolution interpretation of  $s_0(t)$  follows from Brown and Shao (1987) p. 72. Finally an ergodic birth and death process on  $\{0, \dots, d\}$  has distinct eigenvalues  $0, -\lambda_1, \dots, -\lambda_d$  with  $\lambda_i > 0, i=1, \dots, d$  (Keilson (1979) p. 59). It thus satisfies the above conditions, and moreover, the convolution interpretation of  $s_0(t)$  holds.  $\square$

Diaconis and Fill (1989) derived the above expression for  $s_0(n)$  in the case of ergodic stochastically monotone birth and death chains. Their method was based on a construction under which  $s_0(n)$  is a first passage time distribution for a dual birth and death chain. The expression for  $s_0(n)$  then follows from the same Brown-Shao approach as used here. Several examples are presented in their paper in which the eigenvalues are known and  $s_0(n)$  explicitly computed. This includes the symmetric random walk on the cube, where they derive a special case of (5.2).

An example of a skip free to the right, non birth and death chain, satisfying the conditions of Lemma 5.1 is now given:

$$P = \begin{pmatrix} \frac{2}{5} & \frac{3}{5} & 0 \\ \frac{1}{5} & \frac{2}{5} & \frac{2}{5} \\ \frac{1}{8} & \frac{3}{8} & \frac{1}{2} \end{pmatrix}; \quad \tilde{P} = \begin{pmatrix} \frac{2}{5} & \frac{2}{5} & \frac{1}{5} \\ \frac{3}{10} & \frac{2}{5} & \frac{3}{10} \\ 0 & \frac{1}{2} & \frac{1}{2} \end{pmatrix}$$

$$\text{Here } \varepsilon_1 = \frac{3+\sqrt{13}}{20} = .3308, \quad \varepsilon_2 = \frac{3-\sqrt{13}}{20} = -.0303$$

$$s_0(n) = 2.8575(.3308)^n - 1.8575(.0303)^n.$$

In Section 5.3.2 we consider another partial ordering which applies to birth and death chains.

### 5.3. Other partial orderings.

In Section 5.1 we used chain distance partial ordering, and in 5.2 the usual total ordering. We now discuss two partial orderings. Undoubtedly there are many others that can conveniently apply in specific cases.

5.3.1. Consider the following Markov transition matrix, with states 0,1,2,3:

$$P = \begin{pmatrix} .2 & .3 & .5 & 0 \\ .3 & .4 & .2 & .1 \\ .5 & .2 & .15 & .15 \\ 0 & .1 & .15 & .75 \end{pmatrix}$$

$P$  is not stochastically monotone with respect to any total ordering on  $\{0,1,2,3\}$ . However, consider the partial ordering  $(0,1,2) \stackrel{*}{<} 3$ , with  $0,1,2$  not comparable to one another. A general upper set consists of  $\{3\} \cup A$ , where  $A$  is any subset (perhaps empty) of  $\{0,1,2\}$ .  $P$  is stochastically monotone, with respect to this partial ordering if and only if  $P(3,j) \leq P(i,j)$  for  $i,j=0,1,2$ . This is satisfied for the above chain, which is symmetric, thus  $\tilde{P} = P$  is stochastically monotone with respect to  $\stackrel{*}{\leq}$ . Moreover  $P$  is also doubly stochastic, thus  $\pi$  is uniform. An initial distribution  $\pi_0$ , satisfies  $\pi_0/\pi$  monotone decreasing with respect to the above partial ordering if and only if  $\pi_0(i) \geq \pi_0(3)$  for  $i=0,1,2$ . It follows from Corollary 4.1 that for every such  $\pi_0$ ,  $s(n) = 1 - 4\pi_n(3)$ , in particular  $s_i(n) = 1 - 4p_n(i,3)$  for  $i=0,1,2$ .

5.3.2. Consider the following birth and death chain, with states  $\{0,1,2,3\}$ :

$$P = \begin{pmatrix} \frac{1}{2} & \frac{1}{2} & 0 & 0 \\ \frac{3}{4} & \frac{1}{4} & 0 & 0 \\ 0 & \frac{1}{4} & \frac{1}{2} & \frac{1}{4} \\ 0 & 0 & \frac{1}{2} & \frac{1}{2} \end{pmatrix}$$

$P$  is not stochastically monotone with respect to  $0 < 1 < 2 < 3$ , because  $P(1,0) > P(0,0)$ , but otherwise it would be. To salvage stochastic monotonicity define a partial order by  $(0,1) \stackrel{*}{<} 2 \stackrel{*}{<} 3$ , with  $0$  and  $1$  not comparable. Then stochastic monotonicity, with respect to  $\stackrel{*}{\leq}$  is easily shown to be equivalent to:

$$P(0,1) \geq P(2,1)$$

$$P(1,1) \geq P(2,1)$$

$$P(2,3) \leq P(3,3)$$

The above birth and death chain satisfies these constraints and is thus stochastically monotone with respect to  $\leq^*$ . It follows that  $s_0(n) = 1 - 4p_n(0,3)$ , to which (5.4) applies. This is the result we would have obtained if the chain were monotone with respect to the usual ordering.

As a bonus, we also conclude that  $s_1(n) = 1 - 4p_n(1,3)$ , which by (8.9) in the Appendix reduces to:

$$s_1(n) = \sum_{j=1}^3 \left( \prod_{r \neq j} \frac{1 - \beta_r}{\beta_j - \beta_r} \right) \left( \frac{\beta_j^{-P(0,0)}}{P(0,1)} \right) \beta_j^n$$

where  $\beta_1, \beta_2, \beta_3$  are the non-zero eigenvalues of  $P$ .

Next, consider the following birth and death matrix:

$$P_1 = \begin{pmatrix} 1/2 & 1/2 & 0 & 0 \\ 1/2 & 1/4 & 1/4 & 0 \\ 0 & 1/4 & 1/2 & 1/4 \\ 0 & 0 & 1/2 & 1/2 \end{pmatrix}$$

Note that  $P_1$  is stochastically monotone relative to  $(0,1) \leq^* 2 \leq^* 3$ , as was  $P$ . It is also stochastically monotone relative to  $(2,3) \leq^* 1 \leq^* 0$ , since  $P(3,2) \geq P(1,2)$ ,  $P(2,2) \geq P(1,2)$  and  $P(1,0) \leq P(0,0)$ . As a result we have:

$$s_0(n) = s_3(n) = \sum_{j=1}^3 \left( \prod_{r \neq j} \frac{1-\beta_r}{\beta_j - \beta_r} \right) \beta_j^n$$

$$s_1(n) = \sum_{j=1}^3 \left( \prod_{r \neq j} \frac{1-\beta_r}{\beta_j - \beta_r} \right) \left( \frac{\beta_j^{-P(0,0)}}{P(0,1)} \right) \beta_j^n$$

$$s_2(n) = \sum_{j=1}^3 \left( \prod_{r \neq j} \frac{1-\beta_r}{\beta_j - \beta_r} \right) \left( \frac{\beta_j^{-P(3,3)}}{P(3,2)} \right) \beta_j^n$$

Generalizing the above to general  $d$ , we obtain:

Lemma 5.2. (i) Consider a discrete time ergodic birth and death Markov chain on  $\{0, \dots, d\}$ , with eigenvalues  $1, \beta_1, \dots, \beta_d$ , satisfying:

$$P(i,1) \geq P(2,1), \quad i=0,1$$

$$P(i,i+1) + P(i+1,i) \leq 1, \quad i=2, \dots, d-1.$$

Then:

$$s_0(n) = \sum_{j=1}^d \left( \prod_{r \neq j} \frac{1-\beta_r}{\beta_j - \beta_r} \right) \beta_j^n$$

$$s_1(n) = \sum_{j=1}^d \left( \prod_{r \neq j} \frac{1-\beta_r}{\beta_j - \beta_r} \right) \left( \frac{\beta_j^{-P(0,0)}}{P(0,1)} \right) \beta_j^n$$

(ii) Consider a continuous time ergodic birth and death Markov chain on  $\{0, \dots, d\}$ , with eigenvalues  $0, -\lambda_1, \dots, -\lambda_d$ , and satisfying  $q_{01} \geq q_{21}$ . Then:

$$s_0(t) = s_d(t) = \sum_{j=1}^d \left( \prod_{r \neq j} \frac{\lambda_r}{\lambda_r - \lambda_j} \right) e^{-\lambda_j t}$$

$$s_1(t) = \sum_{j=1}^d \left( \prod_{r \neq j} \frac{\lambda_r}{\lambda_r - \lambda_j} \right) \left( \frac{q_{01}^{-\lambda_j}}{q_{01}} \right) e^{-\lambda_j t}.$$

Section 6. Derivation of (1.2).

First, consider a continuous time chain with  $p_t(j,j)$  decreasing in  $t$ , for a state  $j$ . Let  $W$  be a random variable with:

$$(6.1) \quad \Pr(W > t) = \frac{p_t(j,j) - \pi(j)}{1 - \pi(j)} .$$

Letting  $\psi_W$  denote the Laplace transform of  $W$ , we have:

$$(6.2) \quad \psi_W(s) = 1 - s \int e^{-st} \Pr(W > t) dt .$$

Furthermore, with  $\psi_{jj}$  the Laplace transform of  $p_t(j,j)$ :

$$(6.3) \quad \int e^{-st} \Pr(W > t) dt = \frac{\psi_{jj}(s) - \pi(j)s^{-1}}{1 - \pi(j)} .$$

Thus, from (6.2) and (6.3):

$$(6.4) \quad s\psi_{jj}(s) = 1 - (1 - \pi(j))\psi_W(s) .$$

Substitute (6.4) into (2.2) to obtain:

$$(6.5) \quad \psi_{\pi,j}(s) = \frac{\pi_j}{1 - (1 - \pi_j)\psi_W(s)} .$$

The right side of (6.5) is the Laplace transform of  $\sum_{i=1}^N W_i$ , which proves (1.2).

In the discrete time case with  $p_n(j,j)$  decreasing, define  $W$  to be an integer valued random variable with  $\Pr(W > n) = \frac{p_n(j,j) - \pi(j)}{1 - \pi(j)}$ . Then apply the above argument, using probability generating functions in place of Laplace transforms. Again,  $T_{\pi,j} \sim \sum_{i=1}^N W_i$ .

(6.6) Remark. For time reversible continuous time Markov chains with finite state space,  $p_t(j,j)$  is stochastically monotone and thus decreasing for all  $j$  (Keilson (1979) p. 34). The discrete time analogue requires that all the eigenvalues be non-negative.

For ergodic Markov chains in discrete or continuous time, taking values in a finite partially ordered set  $(S, \leq^*)$ , with either  $P$  or  $\tilde{P}$  stochastically monotone relative to  $\leq^*$ , we now argue that  $p_t(j,j)$  is decreasing, where  $j$  is a unique maximal or unique minimal state.

Suppose that  $j$  is a unique maximal (minimal) state and that  $P$  is stochastically monotone. Then  $\{j\}$  is an upper (lower) set and  $p_t(j,j) \geq p_t(k,j)$  for all  $k \in S$ , and  $t$ . Then for  $t_1 < t_2$ :

$$p_{t_2}(j,j) = \sum_{k \in S} p_{t_2-t_1}^{(j,k)} p_{t_1}(k,j) \leq p_{t_1}(j,j).$$

Thus  $p_t(j,j)$  is decreasing in  $t$ .

If  $\tilde{P}$  is stochastically monotone, then by the same argument  $\tilde{p}_t(j,j)$  is decreasing in  $t$ , but  $\tilde{p}_t(j,j) = p_t(j,j)$ , thus  $p_t(j,j)$  is decreasing.

### Section 7. Approximate Exponentiality.

The representation  $T_{\pi,j} = \sum_{i=1}^N W_i$  is useful in studying approximate exponentiality.

First we record a result of the author (Brown (1987)) dealing with geometric convolutions:

Lemma 7.1. Suppose that  $Y = \sum_{i=1}^N X_i$ , where  $\{X_i\}$  is an i.i.d. sequence of nonnegative random variables and  $N$  is independent of  $\{X_i\}$  with  $\Pr(N=k) = q^k p$ ,  $k=0,1,\dots$ . Then:

$$\sup_t |\Pr(Y > t) - e^{-t/EY}| \leq p \left( \frac{EX^2}{(EX)^2} \right) = 2q\rho_Y$$

where  $\rho_Y = \frac{EY^2}{2(EY)^2} - 1$ .  $\square$

Applying Lemma 7.1 and (1.2) we find that:

$$(7.1) \quad \sup_t |\Pr(T_{\pi,j} > t) - e^{-t/\alpha_j}| \leq \pi(j) \frac{EW^2}{(EW)^2} = 2(1-\pi(j))\rho_{\pi,j}$$

where  $\alpha_j = ET_{\pi,j}$  and  $\rho_{\pi,j} = \frac{ET_{\pi,j}^2}{2\alpha_j^2} - 1$ .

Inequality (7.1) tells us that if the first two moments of  $T_{\pi,j}$  behave similarly to those of an exponential distribution, then  $T_{\pi,j}$  is approximately exponential.

For time reversible continuous time chains  $T_{\pi,j}$  is completely monotone (Keilson (1979) p. 11). It follows from Brown (1983) p. 422 that:

$$(7.2) \quad \sup_t |\Pr(T_{\pi,j} > t) - e^{-\alpha_j t}| \leq \frac{\rho_{\pi,j}}{\rho_{\pi,j} + 1}$$

$$(7.3) \quad \sup_t |\Pr(T_{\pi,j} > t) - (1 - \pi(j))e^{-t(1 - \pi(j))/\alpha_j}| \leq \frac{\rho_{\pi,j}}{\rho_{\pi,j} + 1} - \pi(j) .$$

Expression (7.3) is used to bound the error of a modified exponential approximation which takes into account the atom of size  $\pi(j)$  at 0. Both (7.2) and (7.3) hold for first passage times to arbitrary (rather than just singleton) sets. The reason for our restricting attention to singleton sets is that (1.2) will provide the means of expressing the bound in terms of  $\tau/\alpha_j$ , where  $\tau$  is the relaxation time. In applications  $\tau$  is easier to numerically approximate, and to bound, than is  $ET_{\pi,j}^2$ .

We now focus on the moments of  $T_{\pi,j}$ . First recall (Keilson (1979) p. 34) the spectral representation of transition probabilities in continuous time, time reversible chains:

$$(7.4) \quad p_t(j,j) = \pi(j) + \sum_{k=1}^m \mu_{kj}^2 e^{-\lambda_k t}$$

where  $0 > -\lambda_1 \geq -\lambda_2 \geq \dots \geq -\lambda_m$  are the eigenvalues of the infinitesimal matrix.

It follows from (7.4) and (6.1) that  $W \sim UE$  with  $U$  and  $E$  independent,  $E$  exponential with mean 1, and  $\Pr(U=j) = \frac{\mu_{kj}^2}{1 - \pi(j)}$ ,  $k=1, \dots, m$ . Thus,  $EW = EU$ , and  $EW^2 = 2EU^2$ . Thus:

$$(7.5) \quad \alpha_j = ET_{\pi,j} = \frac{1 - \pi(j)}{\pi(j)} EU = \frac{1}{\pi(j)} \sum \mu_{kj}^2 \lambda_k^{-1}$$

$$(7.6) \quad \text{Var}[E(T_{\pi,j} | N)] = \text{Var}(NEU) = \frac{1 - \pi(j)}{\pi^2(j)} (EU)^2$$

$$(7.7) \quad E[\text{Var}(T_{\pi,j} | N)] = E[N \text{Var} W] = \frac{1-\pi(j)}{\pi(j)} [2EU^2 - (EU)^2] .$$

From (7.5)-(7.7):

$$(7.8) \quad \text{Var } T_{\pi,j} = \alpha_j^2 + \frac{2}{\pi(j)} \sum \mu_{kj}^2 \lambda_k^{-2} \leq \alpha_j^2 + 2\tau\alpha_j$$

where  $\tau = \max_k (\lambda_k^{-1})$ , the relaxation time. From (7.8):

$$(7.9) \quad \frac{\rho_{\pi,j}}{\rho_{\pi,j}+1} = 1 - \frac{2\alpha_j^2}{ET_{\pi,j}^2} \leq \frac{\tau/\alpha_j}{(\tau/\alpha_j)+1} .$$

We now summarize:

Theorem 7.1. Let  $\{X(t), t \geq 0\}$  be a continuous time, time reversible Markov chain, with finite state space. Then:

$$(i) \quad \sup_t |\Pr(T_{\pi,j} > t) - e^{-t/\alpha_j}| \leq \frac{\tau/\alpha_j}{(\tau/\alpha_j)+1}$$

$$(ii) \quad \sup_t |\Pr(T_{\pi,j} > t) - (1-\pi(j))e^{-(1-\pi(j))t/\alpha_j}| \leq \frac{\tau/\alpha_j}{(\tau/\alpha_j)+1} - \pi(j). \quad \square$$

The difficulty in dealing with discrete time, time reversible chains, is that due to possibly negative eigenvalues  $T_{\pi,j}$  need not be a mixture of geometric distributions. Thus (7.2) and (7.3) are not applicable. Nevertheless, if  $p_n(j,j)$  is decreasing then (7.1) applies. Moreover, as follows from an argument of Aldous ((1989) p. 183):

$$(7.10) \quad \rho_{\pi,j}^{(\text{discrete})} = \rho_{\pi,j}^{(\text{continuous})} - (2\alpha_j)^{-1} .$$

Combining (7.1), (7.9) and (7.10) we easily derive:

Corollary 7.1. Let  $\{X_n, n=0,1,\dots\}$  be a time reversible Markov chain with  $p_n(j,j)$  decreasing. Then:

$$\sup_t |\Pr(T_{\pi,j} > t) - e^{-t/\alpha_j}| \leq (1-\pi(j)) \left(\frac{2\tau-1}{\alpha_j}\right) .$$

where  $1 > \beta_1 \geq \beta_2 \geq \dots \geq \beta_d$  are the eigenvalues of  $P$ , and  $\tau = (1-\beta_1)^{-1}$ .  $\square$

In the case of a discrete time, time reversible Markov chain, with all nonnegative eigenvalues, the methodology of Brown (1983), p. 422, can be applied to derive:

$$(7.11) \quad \sup_n |\Pr(T_{\pi,j} > n) - \left(\frac{\alpha_j}{\alpha_j+1}\right)^n| \leq \frac{\left(\frac{\tau-1}{\alpha_j}\right)}{\left(\frac{\tau}{\alpha_j}\right)+1}$$

Finally we remark that it follows from an argument of Brown (1987) p. 15, and (1.1), that if  $\pi_t(j)$  is increasing then:

$$(7.12) \quad \sup_t |\Pr(T_{\pi_0,j} > t) - e^{-t/ET_{\pi_0,j}}| \leq \frac{EY}{\alpha_j} + \sup_t |\Pr(T_{\pi,j} > t) - e^{-t/\alpha_j}| .$$

Thus, when  $EY$  is small compared to  $\alpha_j$ , approximate exponentiality of  $T_{\pi,j}$  also yields approximate exponentiality of  $T_{\pi_0,j}$ .

Example. Consider a continuous time random walk on the cube,  $\{0,1\}^d$ , with:

$$q(x, x+\delta_i) = \mu_i > 0, \quad x_i = 0$$

$$q(x, x-\delta_i) = \eta_i > 0, \quad x_i = 1$$

The key quantities are:

$$(7.13) \quad \tau = [\min(\mu_i + \eta_i)]^{-1}$$

$$(7.14) \quad \alpha_j = \sum_{\gamma \neq \emptyset} s_{\gamma}^{-1} \left( \prod_{i \in \gamma} \frac{\mu_i}{\eta_i} \right) \left( \prod_{i \in \bar{\gamma}} \frac{\eta_i}{\mu_i} \right)$$

where  $\gamma$  ranges through non-empty subsets of  $\{1, \dots, d\}$ ,  $s_{\gamma} = \sum_{i \in \gamma} (\mu_i + \eta_i)$ , and  $B_j = \{i: j_i = 0\}$ , the zero components of the vector  $j$ .

$$(7.15) \quad \rho_{\pi, j} = \sum_{\gamma \neq \emptyset} s_{\gamma}^{-2} \left( \prod_{i \in \gamma} \frac{\mu_i}{\eta_i} \right) \left( \prod_{i \in \bar{\gamma}} \frac{\eta_i}{\mu_i} \right) / \alpha_j^2$$

$$(7.16) \quad EY = \sum_{k=1}^d (-1)^{k-1} \sum_{A_k} s_{\gamma}^{-1}$$

where  $A_k$  is the collection of subsets of size  $k$  from  $\{1, \dots, d\}$ .

Now, we simplify by letting  $\mu_i = \eta_i \equiv c > 0$ , obtaining the symmetric walk. In Table 1 below we consider the cases  $d=10$  and  $d=20$ .

In column 1 of Table 1 we compute the error bound (7.2) for exponential approximation of  $T_{\pi, j}$ . In column 2, we replace  $\rho = \left( \frac{ET_{\pi, j}^2}{2\alpha_j^2} - 1 \right)$  by its upper bound  $\tau/\alpha_j$ , obtaining the approximation (i) of Theorem (7.1).

We see that the exponential approximation is quite accurate, and the relaxation time simplification gives a usable though quite conservative upper bound.

In column 3 we compute the refined error bound, (7.3), which adjusts the approximating distribution to mimic the known probability,  $\Pr(T_{\pi, j} = 0)$ . In column 4 the  $\rho$  based bound is replaced by the corresponding relaxation time expression (Theorem 7.1, (ii)). It is seen that this approximation is remarkably accurate. Much is lost in using the relaxation time upper bound,

with the error bound over 100 times too large with  $d=20$ . However, even the conservative bound is quite small.

In column 5 we table the error bound in exponential approximation of  $T_{x,x+1}$ , by use of (7.2) and the column 1 error bound for  $T_{\pi,x+1}$  (note that in the symmetric case the error bound for  $T_{\pi,j}$  is independent of  $j$ ). In column 6, we look at the analogous quantity when the column 2 error bound is used with (7.12). Here, since  $EY/\alpha$  is the dominant contribution to error, little is lost in the relaxation time simplification.

Table 1

|      | Exponential approximation of $T_{\pi, j}$ |                                       | Modified exponential approximation |  | Approximation to $T_{x, x+1}$             |   |
|------|---|---------------------------------------|------------------------------------|--|---|---|
|      | $\frac{\rho}{\rho+1}$                     | $\frac{\tau/\alpha}{(\tau/\alpha)+1}$ | $\frac{\rho}{\rho+1} - \pi(j)$     | $\frac{\tau/\alpha}{(\tau/\alpha)+1} - \pi(j)$ | $\frac{EY}{\alpha} + \frac{\rho}{\rho+1}$ | $\frac{EY}{\alpha} + \frac{\tau/\alpha}{(\tau/\alpha)+1}$ |
| d=10 | $1.22 \times 10^{-3}$                     | $4.25 \times 10^{-3}$                 | $2.40 \times 10^{-4}$              | $3.27 \times 10^{-3}$                          | $1.37 \times 10^{-2}$                     | $1.67 \times 10^{-2}$                                     |
| d=20 | $1.03 \times 10^{-6}$                     | $8.90 \times 10^{-6}$                 | $7.27 \times 10^{-8}$              | $8.04 \times 10^{-6}$                          | $3.34 \times 10^{-5}$                     | $4.04 \times 10^{-5}$                                     |

### Section 8. Comments and Additions.

1) Corollary 4.1 justifies the heuristic interpretation of (1.1) given in Section 1, under appropriate conditions. In general,  $\pi_t(j)$  increasing does not imply that  $s(t) = 1 - \frac{\pi_t(j)}{\pi(j)}$ . The following example, which illustrates this point was suggested to me by David Aldous.

Consider a birth and death process on  $\{0,1,2\}$  with  $q_{01} = q_{21} = 2$ ,  $q_{10} = q_{12} = 1$ . Then:

$$p_t(0,1) = \frac{1}{2} (1 - e^{-4t})$$

which is increasing. But from Lemma 5.1:

$$s(t) = 1 - \frac{p_t(0,2)}{\pi(2)} = 1 - (1 - e^{-2t})^2 = e^{-4t}(2e^{2t} - 1).$$

I have found that in some cases (1.1) can be explained in the following way. There exists a distribution  $\pi'$  such that  $T_{\pi',j} \sim T_{\pi,j}$ , and a stopping time  $Y$  with  $X(Y) \sim \pi'$ ,  $X(Y)$  independent of  $Y$ ,  $\Pr(X(t) = j \text{ for some } t < Y) = 0$ , and  $\Pr(Y > t) = 1 - \frac{\pi_t(j)}{\pi(j)}$ . If this holds we can interpret  $Y$  as the waiting time from  $\pi_0$  to  $\pi'$ . I do not know for how wide a class of situations this interpretation applies.

2) From (4.1):

$$(8.1) \quad d(t) \leq (1 - P_{\pi}(\pi(X) \leq \pi_t(X)))s(t).$$

Suppose that the conditions of Theorem 3.2 hold and in addition  $\pi_0 = \delta_i$ , where  $i$  is the unique minimal state ( $i \leq k$  for all  $k \in S$ ). Then by

Remark (6.6),  $p_t(i,i)$  is decreasing, and thus  $p_t(i,i) \geq \pi(i)$  for all  $t$ .

Thus:

$$(8.2) \quad P_{\pi}(\pi(X) \leq \pi_t(X)) \geq \pi(i)$$

and from (8.1) and (8.2):

$$d(t) \leq (1-\pi(i)) \left(1 - \frac{p_t(i,M)}{\pi(M)}\right).$$

This inequality applies to ergodic stochastically monotone birth and death chains on  $\{0, \dots, d\}$  with  $i=0$  and  $M=d$ . It also applies to the non-symmetric walk on the cube discussed in Section (5.1). Here:

$$d_x(n) \leq (1-\pi(x))s_x(n)$$

where  $s_x(n)$  is given by (5.2), and:

$$\pi(x) = \left[ \prod_{i=1}^d (\alpha_i + \beta_i) \right]^{-1} \left( \prod_{x_i=0} \beta_i \right) \left( \prod_{x_i=1} \alpha_i \right).$$

### Appendix.

We will derive the expressions for transition probabilities that were used in Sections 5.2 and 5.3. These are special cases of more general results I plan to present in a forthcoming paper. The theme, began in Brown-Shao (1987), is that one can develop expressions for transition probabilities and first passage time distributions, which depend on eigenvalues, but not explicitly on eigenvectors. Nor are eigenvectors needed to derive these expressions.

Suppose that  $P$ , an  $m \times m$  matrix, is diagonalizable, so that  $P = ADA^{-1}$ , where  $D$  is diagonal with diagonal entries  $d_1, \dots, d_m$ , which are necessarily the eigenvalues of  $P$ . Since  $P^n = AD^nA^{-1}$  for all  $n = 0, 1, \dots$ , it follows that  $f(P) = Af(D)A^{-1}$  for polynomials,  $f$ , where  $f(D)$  is a diagonal matrix with diagonal entries  $f(d_1), \dots, f(d_m)$ . Let  $\beta_1, \dots, \beta_k$  denote the  $1 \leq k \leq m$  distinct eigenvalues of  $P$ . It follows that for polynomials,  $f$  and  $g$ ,  $f(P) = g(P)$  if and only if  $f(\beta_i) = g(\beta_i)$ ,  $i = 1, \dots, k$ . Now, let  $f(x) = x^n$  and  $g(x) = \sum_{i=1}^k \left[ \prod_{r \neq i} \frac{x - \beta_r}{\beta_i - \beta_r} \right] \beta_i^n$ . The polynomial  $g(x)$  is the well known Lagrange interpolation polynomial (Birkhoff and Rota (1969), p. 215). It is a polynomial of degree  $k-1$ , with  $g(\beta_i) = f(\beta_i) = \beta_i^n$ ,  $i = 1, \dots, k$ . It follows that  $g(P) = f(P)$ , thus:

$$(8.1) \quad P^n = \sum_{i=1}^k \left( \prod_{r \neq i} \frac{P - \beta_r I}{\beta_i - \beta_r} \right) \beta_i^n.$$

This result, (8.1), is discussed in Gantmacher (1960) p. 101, and in Dunford and Schwartz (1958) p. 562.

Assume that we have a skip free to the right Markov chain on  $\{0,1,\dots,d\}$  with  $d+1$  distinct eigenvalues  $1, \beta_1, \dots, \beta_d$ , and transition matrix  $P$ . Distinctness of the eigenvalues implies that  $P$  is diagonalizable, and thus that (8.1) holds. Noting that  $p_n(0,d) = 0$ ,  $n=0, \dots, d-1$ , and  $p_d(0,d) = \prod_{i=0}^{d-1} P(i,i+1)$ , we have from (8.1):

$$(8.2) \quad p_n(0,d) = p_d(0,d) \left[ \prod_{i=1}^d \frac{1}{1-\beta_i} - \sum_{i=1}^d \left( \prod_{r \neq i} \frac{1}{\beta_i - \beta_r} \right) \frac{1}{1-\beta_i} \beta_i^n \right].$$

Letting  $n \rightarrow \infty$  in (8.2), we see that if the Markov chain is ergodic then

$$p_d(0,d) = \pi(d) \prod_{i=1}^d (1-\beta_i), \quad \text{thus:}$$

$$(8.3) \quad p_n(0,d) = \pi(d) \left[ 1 - \sum_{i=1}^d \left( \prod_{r \neq i} \frac{1-\beta_r}{1-\beta_i} \right) \beta_i^n \right]$$

$$(8.4) \quad 1 - \frac{p_n(0,d)}{\pi(d)} = \sum_{i=1}^d \left( \prod_{r \neq i} \frac{1-\beta_r}{1-\beta_i} \right) \beta_i^n.$$

Next, consider  $p_n(1,d)$ , noting that  $p_n(1,d) = 0$ ,  $n=0, \dots, d-2$ ,  $p_{d-1}(1,d) = (P(0,1))^{-1} p_d(0,d)$ , and  $p_d(1,d) = (P(0,1))^{-1} \left( \sum_{i=1}^d P(i,i) \right) p_d(0,d)$ . Furthermore  $\sum_{i=1}^d P(i,i) = \text{tr}(P) - P(0,0) = \sum_{i=1}^d \beta_i + P(0,1)$ . Thus:

$$(8.5) \quad p_{d-1}(1,d) = \pi(d) (P(0,1))^{-1} \prod_{i=1}^d (1-\beta_i)$$

$$(8.6) \quad p_d(1,d) = \pi(d) (P(0,1))^{-1} \left[ \sum_{i=1}^d \beta_i + P(0,1) \right] \prod_{i=1}^d (1-\beta_i)$$

From (8.1):

$$\begin{aligned}
 (8.7) \quad p_n(1,d) &= p_d(1,d) \left[ \prod_{i=1}^d \frac{1}{(1-\beta_i)} - \sum_{i=1}^d \left( \prod_{r \neq i} \frac{1}{\beta_i - \beta_r} \right) \frac{\beta_i^n}{1-\beta_i} \right] \\
 &- p_{d-1}(1,d) \left[ \frac{\prod_{i=1}^d \beta_i}{\prod_{i=1}^d (1-\beta_i)} - \sum_{i=1}^k \left[ \frac{(\sum_{j=1}^d \beta_j) + (1-\beta_i)}{\prod_{r \neq i} (\beta_i - \beta_r)} \right] \frac{\beta_i^n}{1-\beta_i} \right].
 \end{aligned}$$

Substituting (8.5) and (8.6) into (8.7) and collecting terms we obtain:

$$(8.8) \quad p_n(1,d) = \pi(d) \left[ 1 - \sum_{i=1}^d \left( \prod_{r \neq i} \frac{1-\beta_r}{\beta_i - \beta_r} \right) \left( \frac{\beta_i^{-P(0,0)}}{P(0,1)} \right) \beta_i^n \right]$$

$$(8.9) \quad 1 - \frac{p_n(1,d)}{\pi(d)} = \sum_{i=1}^d \left( \prod_{r \neq i} \frac{1-\beta_r}{\beta_i - \beta_r} \right) \left( \frac{\beta_i^{-P(0,0)}}{P(0,1)} \right) \beta_i^n.$$

Similar expressions can be derived for  $p_n(j,d)$ ,  $j=2, \dots, d-1$ .

Acknowledgments.

I am grateful to Professor Ethan Akin who pointed out the Dunford-Schwartz reference, and to Professor David Aldous for suggesting the example in Section 8.

### References.

- Aldous, D. (1989). "Hitting times for random walks on vertex-transitive graphs," Math. Proc. Camb. Phil. Soc., 106, p. 179-191.
- Aldous, D. and Diaconis, P. (1987). "Strong uniform times and finite random walks," Advances in Applied Mathematics, 8, 69-87.
- Birkhoff, G. and Rota, G. (1969). Ordinary Differential Equations, Second Edition. Blaisdell Publishing Company, Waltham, MA.
- Brown, M. (1975). "The first passage time distribution for a parallel exponential system with repair." In Reliability and Fault Tree Analysis, R.E. Barlow, J. Fussell and N. Singpurwalla (eds.). Conference volume published by SIAM, Philadelphia.
- Brown, M. (1983). "Approximating IMRL distributions by exponential distributions, with applications to first passage times," The Annals of Probability, Vol. 11, No. 2, 419-427.
- Brown, M. (1984). "On the reliability of repairable systems," Operations Research, Vol. 32, No. 3, 607-615.
- Brown, M. (1987), "Error bounds for exponential approximations of geometric convolutions." To appear in The Annals of Probability.
- Brown, M. and Chaganty, R.C. (1983). "On the first passage time distribution for a class of Markov chains," The Annals of Probability, Vol. 11, No. 4, 1000-1008.
- Brown, M. and Shao, Y. (1987). "Identifying coefficients in the spectral representation for first passage time distributions," Probability in the Engineering and Information Sciences, 1, 69-74.

- Diaconis, P. (1988). Group Representations in Probability and Statistics.  
Institute of Mathematical Statistics Lecture Notes-Monograph Series, Volume  
11. Institute of Mathematical Statistics, Hayward, CA.
- Diaconis, P. and Fill, J.A. (1989). "Strong stationary times via a new form  
of duality." To appear in The Annals of Probability.
- Dunford, N. and Schwartz, J. (1958). Linear Operations, Part I: General  
Theory. Interscience, New York.
- Gantmacher, F.R. (1960). The Theory of Matrices, Volume I. Chelsea  
Publishing Company, New York.
- Keilson, J. (1979). Markov Chain Models - Rarity and Exponentiality.  
Springer-Verlag, New York.
- Shantikumar, J.G. (1988). "DFR property of first-passage times and its  
preservation under geometric compounding." The Annals of Probability,  
Vol. 16, No. 1, 397-406.